A neural circuit model of flexible sensorimotor mapping: learning and forgetting on multiple timescales

Fusi, Stefano ; Asaad, Wael F ; Miller, Earl K ; Wang, Xiao-Jing

DOI: https://doi.org/10.1016/j.neuron.2007.03.017

Published in final edited form as:
Neuron. 2007 April 19; 54(2): 319–333. doi:10.1016/j.neuron.2007.03.017.

A Neural Circuit Model of Flexible Sensori-motor Mapping: Learning and Forgetting on Multiple Timescales

Stefano Fusi[1,2], **Wael F. Asaad**[3,4], **Earl K. Miller**[3], and **Xiao-Jing Wang**[5]

[1]Center for Neurobiology and Behavior, Columbia University College of Physicians and Surgeons, New York NY 10032 [2]Institute of Neuroinformatics, ETH UNI Zurich, Switzerland [3]The Picower Institute for Learning and Memory, RIKEN-MIT Neuroscience Research Center, and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139. [4]Department of Neurosurgery, Massachusetts General Hospital, Boston, MA 02114. [5]Department of Neurobiology, Kavli Institute for Neuroscience, Yale University School of Medicine, 333 Cedar Street, New Haven, CT06520.

Summary

Volitional behavior relies on the brain's ability to remap sensory flow to motor programs whenever demanded by a changed behavioral context. To investigate the circuit basis of such flexible behavior, we have developed a biophysically-based decision-making network model of spiking neurons for arbitrary sensorimotor mapping. The model quantitatively reproduces behavioral and prefrontal single-cell data from an experiment in which monkeys learn visuo-motor associations that are reversed unpredictably from time to time. We show that when synaptic modifications occur on multiple timescales, the model behavior becomes flexible only when needed: slow components of learning usually dominate the decision process. However, if behavioral contexts change frequently enough, fast components of plasticity take over, and the behavior exhibits a quick forget-and-learn pattern. This model prediction is confirmed by monkey data. Therefore, our work reveals a scenario for conditional associative learning that is distinct from instant switching between sets of well established sensorimotor associations.

Introduction

In simple reflex, a stimulus automatically triggers a stereotyped motor response in a one-to-one fashion. By contrast, adaptive behavior critically depends on the brain's ability to flexibly choose an appropriate response which can vary depending on the specific behavioral context. For example when we see a crosswalk and intend to cross the road, we need to first look left in the US, and right in the UK. The same visual stimulus (the crosswalk) should lead to two different motor responses (look left or look right) depending on the context. If we grew up in the US and we travel to UK for a trip, we can certainly learn to associate to a crosswalk a different motor response. Interestingly we can also retain our bias to look left, as a result of a lifetime practice, and when we go back to US we can immediately remember that bias. This ability indicates that there are probably learning mechanisms operating on multiple timescales: fast components would allow to adapt quickly to new environments, while slow components

Correspondence: X. J.W. (xjwang@yale.edu).

would retain the memory of our experiences on longer times scales. The existence of multiple learning components plays a fundamental role in the decision process and in the learning strategy. In a stable environment, we are requested to respond consistently to sensory stimuli over long time scales and we need to ignore exceptions. For example we do not want to modify the association crosswalk-look left if we live in the US and for some temporary works we need to look right. On the other hand, if we move back and forth between US and UK, we need to adapt to new environments frequently and quickly.

We studied this kind of adaptive behavior by investigating a specific type of flexibility in a controlled laboratory environment. In an oculomotor paradigm (Asaad et al., 1998), monkeys were trained to associate visual stimuli (pictures) with delayed saccadic movements (left or right) (Fig. 1A). The neural mechanism underlying learning has been investigated by recording from prefrontal cortex. Clinical (Petrides, 1985), lesion (Petrides, 1982; Passingham, 1993; Murray et al., 2000), single-unit physiology with behaving primates (Passingham, 1993; Chen and Wise, 1995a; Chen and Wise, 1995b; Chen and Wise, 1996; White and Wise, 1999; Asaad et al., 1998; Asaad et al., 2000), and imaging (Boettiger et al., 2005) studies have shown that the frontal lobes are critical for learning context-dependent ('conditional') visuo motor mapping in humans and nonhuman primates. In order to study the process of learning, the rewarded associations were changed at unpredictable times, and the monkeys had to learn by trial and error a new set of rewarded responses to the same visual stimuli. In particular two visual stimuli (A and B) were initially associated with Left and Right saccadic responses (L and R) respectively. From time to time the associations were reversed (from AL and BR to AR and BL, and vice versa) without any warning to the animal. When the associations were reversed, the monkeys quickly forgot the old associations and then learned the new ones. Interestingly, after a reversal, the animals almost immediately reverted to a chance level performance, followed by learning the new associations in 15-20 trials on average (Fig. 1B). This feature was observed across seven monkeys trained on this paradigm by different investigators (Asaad et al., 1998; Pasupathy and Miller, 2005; Machon et al., 2006). Two other visual stimuli (C and D) were consistently associated with a fixed motor response throughout the experiment. These non reversing stimuli were randomly intermixed with the first two stimuli A and B.

Motivated by this experiment, we have built a biologically plausible decision-making neural circuit model for arbitrary sensori motor mapping. The model is constructed based on the observation that in the experiment of (Asaad et al., 1998) many recorded neurons in the prefrontal cortex responded selectively to the planned motor response. Moreover, the selectivity appeared progressively earlier within each successive trial as the animal learns a correct cue-response mapping, suggesting a role for these cells in learning arbitrary visuo-motor associations. These observations, and others (Chen and Wise, 1995a; Chen and Wise, 1995b; Chen and Wise, 1996), revealed neural correlates of oculomotor reversal learning, but are not sufficient for establishing synaptic mechanisms that causally link the observed neural activity and learning behavior. In the model that we propose, the interplay between decision-making circuit dynamics and reward-dependent synaptic plasticity quantitatively captures the process of learning and forgetting visuo motor associations, both for the observed behavioral data and the single-cell data recorded in prefrontal cortex in the monkey experiment. Importantly, the model gives rise to a surprising prediction, namely behavioral performance is rapidly reset to chance-level by a single error even after the learning process has reached a steady state, which is confirmed by analysis of monkey data.

Our neural circuit model has yield insights into key questions about flexible sensori-motor behavior such as: what determines when we should be flexible and when not to? How much flexible should we be in different situations? Is there a general neural mechanism underlying flexibility? We will show that the ability to decide when to be flexible (i.e. quickly forget and

relearn) emerges naturally from the competition between learning processes which operate on multiple timescales. The learning components which predominantly drive the decision process are selected by the temporal statistics of the environment (how often do the associations change? is there any bias?). Moreover, random behavior is often observed in monkeys when the environment changes (Asaad et al., 1998; Mansouri and Tanaka, 2002). We show that the degree of randomness reflects the extent to which alternative associations are equally likely on a long timescale.

Results

Experimental observations to build the model

Our model is based on two experimental observations. First, our analysis shows that the learning process is roughly independent for each of the two stimuli (visual cue A and B). Indeed, the proportion of correct responses to a stimulus immediately after a reversal is largely unaffected by the prior presentation of the other reversed object. In other words, having seen one object to be reversed does not greatly influence performance on the other. Specifically, the proportion of correct responses to a particular stimulus immediately after a reversal is 0.08 ±0.08 when not preceded by presentations of the other object. If the other reversed object had been presented, this proportion rises only slightly (0.19±0.07 for the first sessions and 0.25 ±0.08 for the last sessions). In what follows we will assume that the interference between the patterns is negligible and we will show that a model based on this hypothesis can reproduce the behavioral data. Secondly, we focus on modeling neurons that are inherently selective to planned saccadic directions, which constituted 29% of the recorded cells in prefrontal cortex. For example, such a cell would fire at a high rate to a stimulus which signals a response in its preferred direction (e.g. Left), and it would fire at a lower rate to all those stimuli which signal a response in the non-preferred direction (Right). Thus, this cell keeps responding according to the chosen motor response (left saccade) which may be associated with stimulus A or B in different blocks of trials. After a reversal, the time at which selectivity becomes apparent gradually moves backward (from the end of trial, to soon after stimulus presentation and early in the delay period), as the monkey learned the new associations. Based on this finding, we propose a scenario in which cue-response associations are learned through the synaptic plasticity of afferents from sensory neurons to response selective cells in a decision circuit.

The decision making model network

The core of the model is a decision making network of integrate and fire neurons with realistic recurrent synaptic excitation (mediated by AMPA and slow NMDA receptors) and inhibition (mediated by $GABA_A$ receptors) as in (Wang, 2002) (see Figure 2A for a schematic representation of the network architecture). Two subpopulations of excitatory neurons represent the direction selective neurons observed in the experiment, which we assume contribute to the selection of the monkey's intended motor response. In our model, when a visual stimulus is presented, the two populations compete, the winner (Left or Right) signals the decision of the model system. When the inputs activated by a visual stimulus are the same for the two neural populations, the network chooses randomly one of the two saccadic movements with equal probability. Any input imbalance would bias the decision. From direct simulations of the spiking neural network model over many trials, we computed the probability that one of the two responses (say Left) is chosen as a function of $\boldsymbol{g}_{Left}$ - $\boldsymbol{g}_{Right}$, the difference in the total synaptic conductances of external inputs to the two neural populations (Fig. 2B). As the difference increases in favor of Left population, the probability to choose Left increases, and eventually the model's behavior becomes deterministic. Notice that the probability depends only on the difference between the two synaptic conductances, and not on the absolute value of the individual conductances ($\boldsymbol{g}_{Left}$ and $\boldsymbol{g}_{Right}$). The 'psychometric function' is well described by a sigmoidal function (solid curve in Fig. 2B). The network behavior with two

different g_{Left}, g_{Right} values (marked by C and D) is shown in Fig. 2C,D for 14 individual trials, with the raster plots from a single model cell selective to Left, and the Left population firing rate. In the first case (C) g_{Left} is larger than g_{Right}, and the decision network chooses Left in about 90% of the trials. In the second case (D) the input synapses are balanced ($g_{Left} = g_{Right}$) and the Left neural pool wins in half of the trials. In each trial, a single stimulus triggers firing activities in both the Left and Right cell populations. Initially they increase together, until the recurrent synaptic input is strong enough that the two neural population firing rates start to diverge from each other. The recruited synaptic inhibition leads to a winner take all competition, so that one population wins and the other one loses. The time at which it is possible to discriminate between the winner and the loser is essentially the point of no return, at which the model system has already made a decision. This time was estimated by computing the latency to the half maximal direction selectivity as follows. The direction selectivity is defined as the relative difference in the firing rate $(r_A\ r_B)/(r_A+r_B)$, similar to the selectivity index used in the experiment (Asaad et al., 1998). Initially the network is in a symmetrical spontaneous state so the selectivity is zero. At the end of a simulated trial, the network has made a decision and the activity is high for one motor response and low for the other, so the direction selectivity is maximal. There is an intermediate time at which the selectivity is half of the final, maximal selectivity, which we define as the latency to half-maximum selectivity. In the simulations it is clear that with a stronger bias (Fig. 2C) the decision occurs earlier, thus the latency is shorter.

It is computationally costly to simulate the full spiking neuron network model for learning process over hundreds of trials. Since the network's decision behavior is well characterized by the choice probability as a sigmoid (softmax) function of the input difference (Fig. 2B), we can use the latter instead of direct simulations when learning is considered. Hence, in any single trial, given g_{Right} and g_{Left}, the network's choice is assumed to be random, with a probability determined by the softmax criterion. The outcome of the stochastic process used to generate the network's decision leads to synaptic changes according to a reward dependent Hebbian learning rule (see the next section). The modified g_{Right} and g_{Left} values are then used to update the decision criterion for the next trial. This procedure generates a variability from session to session that is similar to the one observed in the experiment.

Learning cue-response associations

Learning is modelled by synaptic plasticity from the stimulus selective inputs to the two competing decision neural populations (g_{Left} and g_{Right}). For the sake of simplicity we will first describe this model with a single learning component, its behavior and experimental predictions. As we will show, this will naturally lead us to introduce learning with multiple components (in particular on at least one additional timescale), in order to account for robust and flexible behavior. Given that in the experiment what the monkey learns about one stimulus does not affect the response to the other, we study separately the external input conductances corresponding to different stimuli. In other words, we focus on a single stimulus (say picture A), and consider which of the two responses is triggered by it. We introduce learning rules for strengthening and weakening the synaptic conductances from that input to the two decision neural pools, depending on whether the triggered response is rewarded or not. Each of the two input synaptic conductances is restricted to a fixed range, reflecting the fact that synapses are bounded and the neural activity varies in a limited range. This restriction makes the memory forgetful (Parisi, 1986; Amit and Fusi, 1994; Fusi, 2002), i.e. the mnemonic trace of the past experiences decays exponentially with their age and old visuo motor associations are forgotten. The learning algorithm is schematically summarized in Figure 3A. When the selected response corresponds to the correct association (e.g. AL) and is rewarded, the external inputs to the winning decision neurons are strengthened, whereas those to the losing neurons are weakened. When the selected response corresponds to the wrong association (AR), there is no reward and both external inputs to the two decision cell populations are depressed quickly and brought

towards their minimal values which are assumed to be equal for left and right. The updating rule in the absence of reward is not uniquely determined by the data and other rules are possible (see Experimental Procedures). Learning rate in non-rewarded trials must be much higher (about 20 times) than that in rewarded trials, in order to reproduce the strong reset observed in the behaving monkey after reversal. A simulation of the learning process is shown in Fig. 3B where we calibrated the parameters by fitting the model to the behavioral data (see below). For each trial, given the synaptic inputs for a specific cue, the response of the monkey is decided randomly with the choice probability of Fig. 2B. Then, the input synaptic conductances are updated according to the rules of Fig. 3A, their time courses across several learning reversals are shown in Figure 3B (upper panel). The corresponding probability of choosing one of the two saccades (Left) is shown in the mid panel. The associations are reversed every 60 70 trials. Before each reversal, the performance is high, and the input synapse to the population representing the correct choice is close to its maximal value. After the first error following reversal, both the inputs are reset to the minimal value and the model starts responding randomly, with a performance at chance level. The model usually spends a few trials in such a situation, because each error resets again both synaptic inputs. This is reflected by small fluctuations of the synaptic inputs around zero (upper panel). As the probability of correct responses surpasses a critical threshold, resets due to mistakes become unlikely, the system learns slowly to have more confidence and eventually responds correctly consistently (middle panel). During this process, the time it takes to make a decision, expressed as the latency to the half maximum direction selectivity of neural activity in a trial, is initially long (about 800ms), and becomes considerably shorter (300ms) after the new associations are established (bottom panel).

Model versus Experimental Data

Fig. 4A shows the learning curve after reversal from our model (solid line), superimposed with monkey's behavioral data (open circles) (see Experimental Procedures for fitting the model parameters). In the model, after a learning reversal the chance level performance results from random decisions driven purely by noise (with $g_{Right} \sim g_{Left}$), with no bias for one motor response or another. However, such behavioral data could have different interpretations, since 50%-50% performance can be produced either by truly random choices, or by a strongly biased perseverant behavior of the monkey (e.g. if the monkey responds always left to both stimuli after one mistake). We examined these possibilities by data analysis, our results show that monkey's behavior does not exhibit a bias for one motor response or another, consistent with our model (Supplemental material, Fig. 1S). Moreover, the learning curve in Fig. 4A was obtained by averaging across blocks of trials. Conceivably such a smooth learning curve could arise even if in single blocks, monkey's learning is not gradual but exhibits a sudden transition from a poor to a high performance level, provided that the transition time is random across blocks of trials. We considered this possibility, and found that raw behavioral data do not show obvious evidence for switch-like abrupt transitions after reversal in individual blocks (Supplemental material, Fig. 2S), although statistically it is diffcult to exclude this possibility entirely. Therefore, our model assumption about gradual learning after reversal appears to be compatible with the monkey data.

We found that in order to replicate the fast reset to chance level performance after association reversal, observed in the monkey experiment (Fig. 4A), in the model both synapses g_{Right} and g_{Left} must undergo strong depression when a response choice is incorrect and yields no reward. Although this seems natural intuitively, it gives rise to a specific model prediction, namely that even after learning has reached a steady state the behavioral performance remains very sensitive to the occurrence of any error. We have tested this conclusion quantitatively in several ways. First, we checked that the performance steadily improves with consecutive correct trials (Fig. 4B). About 7 consecutive correct trials are su cient to reach the maximal performance, and

only 3 consecutive correct trials are enough to get to 80% performance (this is compatible with the behavior observed in (Brasted and Wise, 2004)). On the other hand, the performance is reset to chance level after a single error trial, regardless of the number of consecutive correct trials that precede the mistake (Fig. 4C). This strong and unexpected model prediction is thus confirmed by our analysis of monkey's data.

Furthermore, in our model the learning rates are assumed to be independent of the previous history and hence every mistake caused by the wrong decision of the model network should reset the performance (shown in Figure 4C), independently from the fact that the error is due to reversal or to some other reason. Thus, another model prediction is that the performance curve following every mistake (no matter when it occurs) should be similar to the one obtained following the first error after reversal. In Fig. 4D we show that our prediction is confirmed by additional data analysis. The performance after every mistake is plotted for the experiment (empty circles) and for the model (solid line). The model and the data learning curves match surprisingly well, indicating that our prediction was correct. We also considered the time course of the neural activity in the decision network model during associative learning. Following the first mistake after the reversal, the synaptic inputs are reset to their minimum value and hence the decision is slow (see Fig. 2D). The time to the half of the maximum selectivity is the longest. As learning proceeds, biased synaptic inputs lead to faster firing dynamics in decision neurons (see Fig. 2C), so that the decision time is progressively shortened. Using direct simulations of spiking neuron network model, we found that our model was able to reproduce the electrophysiological data of neural latency to half maximum selectivity in the experiment of reversal learning (Fig. 5), notably with the same model parameters calibrated only by behavioral data.

Learning to respond probabilistically

To achieve random decision making after reversal requires fine tuning of model parameters, because the choice probability is very sensitive to small differences in the two inputs (see Fig. 2B): the range over which the decision behavior is stochastic is small compared to the total synaptic conductances of external stimulation $(g_{Left} - g_{Right})/(g_{Left} + g_{Right}) \sim 3\%$. Therefore any heterogeneity (e.g. random connectivity) can disrupt the mechanism underlying the stochasticity of choice behavior. Here we suggest that this fine tuning can be accomplished in the nervous system by learning on a long timescale. The idea is that, if we add a slower component of learning (using a similar algorithm as for the fast component, see Experimental Procedures, but with smaller learning rates), the memory window within which experiences are kept in memory can be extended to span several blocks of trials in which the same stimulus is remembered to be associated sometimes with one motor response, sometimes with the other (Fig. 6A). If the overall fraction of trials is the same for each of the two responses to be correct, the integration over blocks of trials of the slow components tends to create a symmetric input configuration which makes the two responses equally probable. This is shown in Fig. 6B where the slow components of plastic synapses $g_{Left,Slow}$ and $g_{Right,Slow}$ are initially different (in favor of choosing Left), but become equal through learning across several blocks of trials. At the same time, the fast synaptic components allow the system to learn the correct associations within each block. Every mistake resets the fast components, bringing the system back to the balanced configuration determined by the slow components. This scenario is supported by the experimental observation that, in the early stages of training, monkeys very often adopt strong biases to respond in only one direction. Only once their training is more advanced do they lose these biases and the two choices became equally likely. The experiment has actually been designed to avoid any bias in the response: left and right saccades in response to one specific stimulus are rewarded in the same proportion of trials across many blocks. Our model with dual timescales of synaptic plasticity provides a candidate mechanism for monkey's slow learning through the training process. In fact, with the slow learning component, random

behavior simply reflects the long term statistics of the visuo motor associations across many blocks. In the specific case of the experiment which we considered, the reset to the chance level reflects the 50% 50% statistics of rewarding left and right saccades for all the stimuli whose associations are reversed. The ability of encoding this balanced probability of reward does not depend on the specific learning parameters. Moreover, any built in bias in the network can be compensated by the long timescale learning mechanism. For example, if the number of a erents to the neurons selective for Left is larger than the number of a erents to the Right neural pool, the network would tend to have a marked preference for choosing Left, after an error driven reset of the fast synaptic components. Because Left is chosen excessively even in blocks of trials when the correct and rewarded response is Right, the learning process leads to a gradual depression of $g_{Left,Slow}$ and potentiation of $g_{Right,Slow}$ (Fig. 7). Eventually, the slow synaptic components would compensate for the bias in such a way that the synaptic inputs to the two populations become nearly equal. Therefore, after a reset of the fast synaptic components the network restores to random behavior with a performance at chance level (Fig. 7).

In general the slow components approximately encode the reward history, i.e. the probability that a particular motor response is rewarded on a long timescale (Fig. 8A-D). We found that in our model the overall probability of choosing a motor response matches the probability of rewards for that choice, when averaged across blocks of trials (Fig. 8D). For example, if the blocks in which left is rewarded in response to stimulus A are longer than the blocks in which Right is rewarded, then the slow components will bias the response in favor of Left. After a mistake, the probability of choosing left would be higher than the probability of choosing right (Fig. 8B). If the associations are never reversed, then a single mistake should not lead to random behavior because the slow components will consistently bias only one motor response. This model prediction is confirmed by analyzing the experimental data (Asaad et al., 1998) for the two stimuli whose associations were never reversed (Fig. 8E). The e ect is striking: for the non reversing stimuli one error does not compromise the performance of the monkey. The next time the same cue is presented again, the monkey responds almost always correctly, in contrast to what happens for the reversing stimuli, for which a single mistake leads to chance level performance.

Discussion

A neural circuit model of arbitrary sensori-motor mapping

In summary, we proposed a spiking neuron network model endowed with reward dependent plasticity for learning arbitrary sensori motor associations. Unlike more abstract ‘cognitive type’ models, such biophysically based modeling is necessary for mechanistically explaining the observed behavior in terms of the underlying cellular and synaptic events. Our plasticity rule for the stimulus dependent synaptic conductances is consistent with existing reinforcement models (Sutton and Barto, 1998; Williams, 1992). The stochastic Hebbian learning rule was introduced in (Amit and Fusi, 1994; Fusi, 2002), in which each synapse is binary and it is potentiated or depressed with probabilities. We showed that, with the addition of reward dependence (see also (Soltani and Wang, 2006)), this learning rule in a decision circuit is suitable for describing flexible sensori motor learning. We would like to emphasize that this model not only reproduced previously reported experimental data, but more importantly has yielded the unexpected prediction of fast resetting to random performance by single errors even after the learning process has reached a steady state. We have put this model prediction to test in multiple ways, each time it was confirmed by data analysis of the monkey experiment (Fig. 4). The results from model simulations and behavioral data analyses reported here collectively suggest a novel and specific scenario for conditional associative learning, in contrast to a different scenario in which the system does not learn after reversal, but switches between pre

wired neural representations depending on different contexts (in the case of the experiment: AL-BR, AR-BL) (Deco and Rolls, 2003; Deco and Rolls, 2005; Salinas, 2004).

How to be flexible: when and by how much

When the environment changes we often need to modify our behavior and respond in a different way to some stimuli. When we move between two or more environments, we can flexibly adapt either by erasing the old sensori motor associations and by learning the new ones, or by switching to a previously memorized set of sensori motor associations which guarantee reward in the new environment. In our work we studied the first type of flexibility, which is also a widely observed behavior in monkey experiments, and it is certainly the initial behavior also in the case in which the animal eventually adopts a switching strategy (see below for a more about the second strategy). The processes of forgetting and learning occur at a certain rate, which can be modulated based on the past experience to adapt more rapidly to new environments when we know that the environment changes. Normally we do not want to modify our set of visuo motor associations too rapidly, because in a stable environment, exceptions should be ignored. However if the environment changes often enough, it becomes rewarding to forget quickly. Our work suggests that flexible sensori motor mapping can be conceptualized in terms of synaptic plasticity of sensory inputs to a decision network responsible for action selection. In our model and in the observed monkey behavior, the old associations are quickly forgotten to make room for the new ones only for those stimuli whose associations are reversed from time to time. Interestingly, for these stimuli, the old associations are practically reset after every single error. Instead, for those stimuli which require a consistent response over long time scales, no fast reset is observed. So the behavior of the monkey reflects the statistics of the sensori motor associations on multiple timescales. Indeed for the stimuli whose associations are reversed, the rewarded responses are consistent for tens of trials (i.e. the duration of the blocks in which the responses are not reversed), whereas for the other stimuli the rewarded responses are consistent across thousands of trials in the entire experiment.

What is the neural mechanism underlying flexibility?

The flexible behavior described in the previous section emerges naturally by introducing learning on multiple timescales: for those stimuli whose associations are reversed, the slow components which bias the response reflect the statistics across many different blocks, and they are balanced because the two motor responses, left and right, are rewarded in an equal number of cases. When the slow synaptic components are not biased, they essentially do not play any role in the competition between the two decisions corresponding to the two motor responses left and right, and the fast components can dominate. The reset after one mistake is an expression of a fast process, and it is predicted to be observed only when the slow components are balanced. In the case in which there is a preference for either left or right on long timescales, then the model predicted that there should be no reset, because both the fast and slow components are biased towards one response. Such a prediction has been verified in the behavioral data. Therefore, the same model, with the same parameters, reproduces both the reset behavior of the balanced case of reversing stimuli and the no reset behavior of non reversing stimuli. What determines the difference between the two behaviors is only the statistics of the visuo motor associations on multiple timescales. Notice that our decision making model based on the competition between two populations of neurons representing the two motor responses provides a simple way of selecting the dominant bias: the final response depends on the difference between the synaptic inputs to the two populations, so balanced components like the slow ones for the reversing stimuli, do not contribute at all to the competition between the neural populations encoding the alternative motor responses. It is worth noting that although we focused on a task with two possible responses, our model can be readily generalized to situations with a larger repertoire of motor outputs, such as

associations between multiple sensory stimuli with four saccadic movements (Chen and Wise, 1995a; Chen and Wise, 1995b; Chen and Wise, 1996).

Forget-and-learn versus instant switch

The switch strategy is intuitively appealing and appears commonplace in human behaviors. However, evidence is scarce that monkeys can learn new conditional sensori motor associations and then reverse them instantaneously. As mentioned earlier, instead of switching, rapid resetting and slower relearning have been consistently observed in seven monkeys in our experiment (Asaad et al., 1998; Pasupathy and Miller, 2005; Machon et al., 2006). A possible explanation lies in the fact that there were non reversing stimuli which were randomly intermixed with the two reversing stimuli (A and B), hence it was not obvious for monkeys to adopt the strategy of simply switching from one cue response mapping to another after each reversal. However, fast reset and slower relearning have also been observed in other experiments. In a similar visuo motor experiment, Chen and Wise (1995a) focused on acquisition of novel conditional associations but occasionally tested reversals. They reported that after a reversal, '*The monkeys usually repeated the response learned in the preceding block of trials, then switched to a trial-and-error strategy, and eventually learnt the altered instructional significance of the stimulus*'. In another experiment in which cue response-reward contingencies were changed from one block of trials to another, after a reversal monkeys reattained high performance in 10-20 trials (Fig. 2B in (Matsumoto et al., 2003)). These observations are consistent with our model.

This is not to say that, depending on the task type and design, or how long the animals are trained, animals cannot adopt the switch strategy. For instance, simpler reversal learning of cue reward contingencies can be very fast, within a few trials (Kennerley et al., 2006). Also, switching between behavioral contexts has been observed when explicit cues were used to signal which rule was currently in e ect (Wallis et al., 2001). In arbitrary sensori motor mapping tasks, switch strategy is more likely if all stimuli are remapped at the same time. This, however, is not a typical situation in real life. In our experiment, by design we used non reversing stimuli which were randomly intermixed with the two reversing stimuli (A and B), hence it was not obvious for monkeys to adopt the switching strategy. Instead, monkeys showed a behavioral pattern of learning, forgetting and relearning, not instant reversals. This allowed us to observe multiple episodes of associative learning during each recording session. It is conceivable that after a longer training period, monkeys could eventually show the switch strategy. If so, at that stage, the behavior would in a sense become more stereotyped, not suitable for studying the dynamical process of flexible associative learning.

Yu and Dayan (2005) argued that switching strategy is desirable only when errors are most likely to be caused by unpredictable change of cue-response contingencies (unexpected uncertainty), not by unreliability of cue response relationship within a block of trials (expected uncertainty). They hypothesized the existence of two neuromodulator systems, related to acetylcholine and norepinephrine, that signal expected and unexpected uncertainty respectively. In our experiment, cue-response contingencies change in an unpredictable way (reversals occur at random times), generating unexpected uncertainty. The animal can become aware of this type of uncertainty only if it can retain memory across blocks of trials in which the contingencies are different, and hence slow learning components are needed both in our model and in the Yu-Dayan model and they play a similar role. On the other hand, in the experiment we modeled there is no obvious expected uncertainty as the correct response is always unambiguous.

Interestingly, in the experiment that we analyzed, most of the errors lead to a fast reset of the associations. However, in order to fit the model to the data, we needed to assume that in a small fraction of cases, 7%, no synaptic modification followed the erroneous response (see the

Experimental Procedures for more details). The simplest explanation of these exceptions would probably be that synapses are not updated when no reward is expected in this small fraction of trials, for reasons that are unknown. For instance, in such a trial the monkey could be distracted or simply tired, hence was aware that it was going to make a mistake and did not expect to get reward. This view is consistent with the fact that the errors not leading to a reset occurred seemingly at random times. An alternative explanation might be related to some form of internal estimate of expected uncertainty: the monkey knew that a correct response was very likely to lead to a reward, but it might not be completely certain because sometimes it made mistakes for reasons which seemed not to be under control, but due to some external unpredictability. If so, one would expect that with an increased expected uncertainty, the fraction of erroneous responses not leading to a reset would be higher. Such a behavior would make the system more robust to expected uncertainty. Manipulation of expected uncertainty in sensori motor mapping tasks in future work would shed insights into this issue.

Note that whereas Yu and Dayan (2005) used a Bayesian inference approach to understand the relation of two neuromodulators to expected and unexpected uncertainty, our model consists of a biophysically based circuit of spiking neurons which allowed us to capture both behavioral and single unit physiological data quantitatively, and to probe mechanistic questions about synaptic plasticity underlying flexible associative learning. Moreover, Our model can be extended to exhibit a switching strategy, based on the idea that alternative contexts are represented internally as coexisting and competing attractor states (Curti et al., 2006). This interesting topic is beyond the scope of the present paper, and will be pursued elsewhere.

Random behavior for equally probable alternatives

The introduction of an additional, slow learning component enabled the model to display chance level stochastic decisions robustly, in spite of cellular heterogeneities that tend to bias systematically the network's choice behavior. Random choice behavior after reversal requires the synaptic inputs to the two decision neural pools to be balanced, so that they are perfectly symmetrical. Such a symmetry in principle can be achieved in a neural system in a few ways. One possibility is through homeostatic regulation (Turrigiano, 1999), which e ectively renders a network homogeneous in spite of cellular or synaptic heterogeneities (Renart et al., 2003). Here we suggested that slow synaptic plasticity provides a natural mechanism which harnesses the feedback from the external world.

Furthermore, slow learning induces adaptive synaptic changes that reflect the statistics of the real world. Indeed, we showed that, across many blocks of trials, if the two response options are rewarded with a certain relative proportion of time, then the model's choice probability matches the reward probability when the fast learning components are reset (i.e. after every mistake). In other words, response choices which are determined by the slow learning components are selected in a proportion which matches the relative reinforcement obtained on these choices, thus the model behaves according to the so called 'matching law' (Herrnstein et al., 1997; Sugrue et al., 2004; Corrado et al., 2006; Soltani and Wang, 2006; Loewenstein and Seung, 2006). Matching behavior has been typically studied using foraging type tasks; to our knowledge it has not been reported for conditional associative learning thus represents a prediction of our model. In the experiment of (Asaad et al., 1998), across trial blocks, the reward fraction for the two possible responses is 0.5 for a reversing stimulus, and 1 (or 0) for a non-reversing stimulus. It would be interesting in future experiments to manipulate this reward fraction over a continuous range, e.g. by using different lengths of trial blocks in which the two responses are alternatively rewarded. Our model predicts that monkey's performance after reversal depends in a graded manner on the long term reward fraction of motor responses to a sensory stimulus, according to the matching law. Therefore, the reward history can be

profitably used to guide the decision making process. If confirmed, this would constitute strong evidence in support of the hypothesized slow learning process.

More generally, we expect that the learning rates are not fixed but depend on the task design. This is because temporal statistics determines the relative distribution of the different learning components and hence the effective learning and forgetting rates. For example, in our model, the forgetting rate is much higher for reversing stimuli than for non reversing stimuli, simply as a consequence of their different reward statistics. These considerations change dramatically the perspective of studying learning in psychophysics and in other experiments: learning models cannot ignore what happens on other time scales because at any time scale we choose to study the phenomenon, the learning rates are affected by the statistics on all the other time scales. Here we studied the simple case of two learning rates. However a continuous distribution of time scales would produce similar results and it is probably desirable when the relevant time scales for the task are not known a priori. An alternative explanation, fully compatible with our model, is to assume that the synapses might change their inherent learning rate, as in a recently proposed cascade model (Fusi et al., 2005). Memory performances are higher with modifiable learning rates than with a static distribution of learning rates because each synapse changes adaptively the rules by which it is modified (meta learning) (Schweighofer and Doya, 2003; Soltani et al., 2006). Moreover the dynamic learning rates of the cascade model gives rise to power law forgetting curves observed in many experiments (see e.g. (Wixted and Ebbesen, 1997)).

Large-scale circuit basis of flexible sensori-motor mapping

In order to explore synaptic mechanisms and for the sake of simplicity, we have chosen to consider a biophysically realistic microcircuit model of decision making, e.g. in prefrontal cortex. However, learning flexible sensori-motor mapping is likely to involve a large brain network. In particular, the basal ganglia appears to play a major role, as shown by behavioral (Murray et al., 2000), physiological (Pasupathy and Miller, 2005) and imaging (Tanaka et al., 2004; Boettiger et al., 2005) studies. The medial temporal structures are also important, presumably because of their role in long term memory (Murray et al., 2000; Wirth et al., 2003). This raises the question as to the respective roles of different brain regions in conditional sensori-motor learning. It was recently reported that in the same conditional sensori motor association task, after a reversal, caudate cells show selectivity to the intended motor response earlier than prefrontal cells, suggesting that basal ganglia could be involved in the selection of the choice, especially in the early learning phases after reversal, while prefrontal cortex encodes the correct motor response when the correct associations are established (Pasupathy and Miller, 2005). Previous fMRI (Tanaka et al., 2004) and modeling (Daw et al., 2005) studies suggest that basal ganglia and prefrontal cortex are engaged in signaling rewards at different (short versus long) timescales, it remains unclear whether this view is consistent with the electrophysiological data of (Pasupathy and Miller, 2005) in conditional learning tasks.

In our model the fast components of learning are practically reset after one mistake, and the slow components dominate the selection of the motor response. The fast components might be responsible for the choice selection in prefrontal cortex while the slow components might control the choice operated by the circuit in the basal ganglia. After one mistake, the fast components of prefrontal cortex might be reset and the basal ganglia might take control and generate stereotyped responses in a specific task (left and right responses with the probability of reward observed over many blocks). It remains to be elucidated in future work whether this view is compatible with the data of (Brasted and Wise, 2004).

The cortico striatal pathway is known to exhibit synaptic plasticity that is strongly modulated by dopamine signals (Reynolds et al., 2001; Reynolds and Wickens, 2002); there is also evidence that dopamine influences long-term synaptic plasticity in prefrontal neurons (Otani

et al., 2003; Huang et al., 2004). It would be interesting to see whether either of these synaptic pathways display plasticity at disparate timescales. Future research on these critical issues will help to elucidate the cellular and circuit mechanisms of conditional associative learning.

Experimental procedures

Analysis of behavioral and neural data

The details of the experimental protocol and the recording techniques can be found in (Asaad et al., 1998). The performance P, at any time point in a trial sequence (e.g. kth trial after reversal, k=1,2,...), is estimated for each visual stimulus separately as the number of correct trials over the total number of trials in which a particular stimulus is presented. The total number of trials includes also the incorrect trials in which the monkey does not respect the protocol (e.g. when it breaks the fixation during the delay). The probability of correct response P is plotted with a confidence interval of 68%, given by (Meyer, 1965):

$$P_{\text{up, down}} = Pn + 1/2 \pm \sqrt{\frac{P(1-P)n + 1/4}{n+1}} \quad (1)$$

where n is the total number of trials. The performance of Figure 1B and 4A is the performance across blocks of trials in which the associations were consistent as a function of the number of trials from reversal averaged over all stimuli. The performances of figures 4C,4B,8E represent the proportion of cases in which the monkey responds correctly following a specific sequence of correct and incorrect trials. In particular the performance of Figure 4B is the proportion of correct trials after a sequence of at least one error, followed by n consecutive correct trials. Figures 4C and 8E show the performance which follows a sequence of least one error, n consecutive correct trials, and then one error again. In both figures 4B and 4C,8E, the sequences are identified by considering only the trials in which a particular stimulus is presented.

Decision neural network model

The architecture and the all parameters of the network of 2000 integrate-and-fire neurons (1600 excitatory, and 400 inhibitory) are identical to the one proposed in previous work (Wang, 2002; Brunel and Wang, 2001). Briefly, two neural pools are selective to the saccadic directions Left and Right. Within each pool, there are strong recurrent excitatory connections between pyramidal cells, that underlie slow ramping activity during stimulus presentation. The two neural pools compete with each other through shared feedback inhibition by GABAergic interneurons (see Figure 2). Synaptic connections are modeled as realistic AMPA, NMDA, $GABA_A$ receptor mediated currents. In addition to the recurrent synaptic currents, all neurons receive an external input of the AMPA type, at 2400 Hz (2400 pre-synaptic neurons firing at 1Hz). The external spikes are generated with a Poisson statistics. During the visual stimulation a fraction $\chi_e = 0.01$ of the external inputs to the excitatory neurons, and a fraction $x_i = 0.33\chi_e$ of the external inputs to the inhibitory neurons is driven to 7.9 Hz for 500ms, 100ms after the stimulus onset to mimic the latency of stimulus induced neural signals observed in prefrontal cortex. Following the stimulus, the same fraction of cells fires at 4.6 Hz until the end of the trial. The latency to the half-maximum selectivity is defined as in (Asaad et al., 1998), for both the neural experimental data and the simulations.

Learning dynamics

In order to modulate the external input and generate a bias for one response, the external synapses to the excitatory decision neurons are assumed to be plastic and with two possible synaptic conductances: 2nS when the synapse is depressed, and 3.6nS when it is potentiated. Because we assume that the associations are learnt independently for each stimulus, we can

focus on the synaptic inputs from a single stimulus to the two decision neural pools. The dynamic variables which describe the learning process are the fractions of stimulus-specific synapses (i.e. a fraction x_e of the total excitatory synapses) in the potentiated state, denoted by c_y, where $y = L, R$ indicates the target population (Left or Right). We dropped the index which would denote the visual stimulus as the two stimuli will always be considered separately. Following each stimulus presentation and a chosen response, there is a reward if the association is correct, and no reward otherwise. At the end of each trial, the c variables corresponding to the presented cue are updated according to:

$$c_y \rightarrow c_y + q_+ \left(r, v_y\right)\left(1 - c_y\right) - q_- \left(r, v_y\right) c_y \tag{2}$$

where $q_+(r,V_y)$ is the rate of potentiation and it is a function of whether the choice y results in a reward or not ($r = R$ or NR) and of the activity of the target neural population V_y. V_y has essentially only two values, corresponding to the two possible decisions of the system. We denote these values by H (=high activity) and S (=spontaneous activity). q_- is the rate of depression. Notice that all c variables are in the range [0,1].

Analysis of learning

When the stimulus is presented, the probability of choosing one of the two saccades (say Left) has the following form:

$$P_L = \frac{1}{1 + e^{-(c_L - c_R)/\sigma}} \tag{3}$$

Figure 2 shows that this sigmoidal function with $\sigma = 0.05$ actually matches the probability of choosing left estimated by simulating the full network of integrate-and-fire neurons. For each pair c_L,c_R we can also determine the average latency to the half maximum selectivity by simulating the full network for 100 trials. The average is estimated only over correct trials. This latency expressed in *ms* is well fitted by the following function:

$$T\left(c_L, c_R\right) = 180 + 555 e^{-(c_L - c_R)/\sigma_T} \tag{4}$$

where $\sigma_T = 0.074nS$. This approximation is good when c_R is close to 0, which is certainly true for most of the trials of the learning process that we intend to describe (see below).

The learning process can be described as a iterative dynamics that give rise to a trajectory in the space of c_L and c_R. For each pair c_L,c_R, the probability of choosing left P_L is determined by Eq. 3. Given this probability, the proportion of cases in which the simulated network gets reward can be estimated, and then it can be used to move to the next point of the space c_L,c_R by using Eq.2.

We now make a few preliminary considerations to reduce the independent variables which describe the learning process. Many combinations of c_L,c_R modifications are equivalent in terms of behavior. For example reducing the input to Right or increasing the input to Left has the same effect on P_L, which depends only on the difference between c_L and c_R. We assume that when a target population wins and there is reward, $q_+(R,H) = q_+^R$ is the only independent learning rate and $q_-(R, H)$ is set arbitrarily to zero. Analogously $q_-(R,S) = q_-^R$ and $q_+(R, S) = 0$. For the no reward cases, in order to reproduce the experimentally observed resetting to chance level performance, we clearly need to symmetrize the inputs to the two target

populations as quickly as possible. This means that the synaptic changes must bring both c_L and c_R rapidly to the same equilibrium distribution $c_{L,R} = 1/(1+ q_- (NR, HorS)/q_+ (NR, HorS))$. This distribution represents a sort of a baseline, on top of which the associations are learned. We can reproduce the experimental data with any baseline, so we set it arbitrarily to zero by making the assumption $q_+ (NR, HorS) = 0$ and $q_- (NR, HorS) = q_-^{NR}$. Notice that the data can be reproduced also with other updating rules. For example if we assume that after learning c = 1 for the correct association and c = 0 for the incorrect one (which is what we get for our updating rule in the presence of reward), then we can also have $q_- (NR, H) = q^{NR}_-$ and $q_- (NR, S) = 0$ for the no reward case. This updating rule brings also to a configuration in which the inputs to the two target populations are symmetric and hence it is compatible with the data. The simulations of Figures 3B are done as follows: 1) c_L and c_R are initialized to 0; 2) the probability P_L is computed as given by Equ.3 and the latency to the half maximum selectivity is computed according to Equ.4; 3) Left population is chosen randomly to be the winner (V_L = H,V_R = S) with probability Q_L; 4) depending on the rule in effect, reward is received or not 5) the c_L and c_R are updated according to Equ.2. Points 2 to 5 are repeated for every trial and c_L, c_R, P_L are plotted as a function of the number of trials. The simulations in Figures 6B,7A, 8A,8B,8C are done in the same way, but the whole procedure is repeated 100 times starting from the same initial condition, every time with a different seed for the random process of point 3. The average c_L, c_R and P_L are plotted as a function of the number of trials from the beginning of the simulation. The average latency to half maximum selectivity is plotted as a function of the number of correct trials from reversal in Figure 5. The details about the mean field analysis are reported in the Supplemental Material.

Fast and slow components of learning

The total plastic input is assumed to be made of fast (c_L^f, c_R^f) and slow components (c_L^s, c_R^s). The fast components are the ones previously introduced. The slow components have the same dynamics as the fast components (see Equ. 2), but with smaller learning rates. Moreover, when the choice yields no reward, the input synapses onto the losing neural population (with low firing rates) are potentiated with a learning rate $q_+ (NR, S) = r_+^{NR}$, hence representing a type of anti Hebbian learning for the slow component (r denotes a learning rate of a slow component, whereas q indicates a learning rate of a fast component). They affect the probability of choosing Left as follows:

$$P_L = \frac{1}{1+e^{-\left(\left(p_s c_L^s + p_f c_L^f\right) - \left(p_s c_R^s + p_f c_R^f\right)\right)/\sigma}}$$

where p_s and p_f are the fractions of slow and fast components respectively (in the simulations: $p_s = 0.4$, $p_f = 0.6$). In the Supplemental Material we prove that slow components are balanced when the motor responses are equally rewarded for a wide range of learning parameters.

Fitting the model to the behavioral data

For each set of the learning parameters $\{q_+^R, q_-^R, q_-^{NR}\}$ we compute the performance of the monkey for the three curves in Figure 4A-C. The parameter space is explored with a Monte Carlo which minimizes the χ^2 distance between the neural data and the model points. The confidence intervals of the χ^2 are estimated as described in Equ. 1. The model would always converge to a maximal performance of 100% correct trials. However, monkeys always make mistakes, even after learning (the performance curve saturates at a level below 100%), which are probably due to a number of possible reasons. The mechanisms responsible for these errors are not modeled here, but in order to reproduce the experimental data we need to take them into account. We do it by introducing a randomly chosen fraction of trials f_{err} in which the

decision of the monkey is controlled by an unspecified mechanism. We assume that in these trials the synapses of our network are not updated. The choice probability used to compare to the monkey data is then $P_{L'} = P_L\,(1\text{-}2\,f_{err}) + f_{err}$ where P_L is the performance of our decision making network ($P_L \in [0,1]$). f_{err} is unknown and it is determined by the Monte Carlo. The parameters for the best fit are: $q_+^R=0.021, q_-^R=0.073, q_-^{NR}=0.96, f_{err}=0.071$. With these parameter values, the comparison between the model and the behavioral data (59 data points) are shown in Fig. 4A-C, with $P > 0.36$ in a χ^2 test.

Compensation of a bias due to heterogeneity

The bias β is introduced as:

$$P_L = \frac{1}{1+e^{-\left(\beta\left(p_s c_L^s + p_f c_L^f\right) - \left(p_s c_R^s + p_f c_R^f\right)\right)/\sigma}}$$

where β represents some kind of fixed heterogeneity which does not change during learning. c_R^s and c_L^s can however compensate this bias to produce balanced inputs to Left and Right when the two responses are rewarded with the same probability.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the NIH grant DA016455 (XJW and SF), the NIH-CRCNS grant NS50944 (XJW), the SNF grant PP00A-106556 (SF) and the NINDS grant 5R01NS35145 9 (EKM) and it was partly done when XJW was at Brandeis University and SF was affliated to the Institute of Physiology in Bern (Switzerland), visiting XJW in his lab at Brandeis. The authors are grateful to C.D. Salzman, L.F. Abbott, and N. Brunel for helpful comments on the manuscript and to N. Daw for useful discussions.

References

Amit DJ, Fusi S. Learning in neural networks with material synapses. Neural Comput 1994;6:957–982.

Asaad WF, Rainer G, Miller EK. Neural activity in the primate prefrontal cortex during associative learning. Neuron 1998;21:1399–1407. [PubMed: 9883732]

Asaad WF, Rainer G, Miller EK. Task-specific neural activity in the primate prefrontal cortex. J Neurophysiol 2000;84:451–459. [PubMed: 10899218]

Boettiger A, Charlotte A, D'Esposito M. Frontal networks for learning and executing arbitrary stimulus-response associations. J Neurosci 2005;25:2723–2732. [PubMed: 15758182]

Brasted PJ, Wise SP. Comparison of learning-related neuronal activity in the dorsal premotor cortex and striatum. Eur J Neurosci 2004;19:721–740. [PubMed: 14984423]

Brunel N, Carusi F, Fusi S. Slow stochastic Hebbian learning of classes in recurrent neural networks. Network 1998;9:123–152. [PubMed: 9861982]

Brunel N, Wang XJ. Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. J. Comput. Neurosci 2001;11:63–85. [PubMed: 11524578]

Chen LL, Wise SP. Neuronal activity in the supplementary eye field during acquisition of conditional oculomotor associations. J Neurophysiol 1995a;73:1101–1121. [PubMed: 7608758]

Chen LL, Wise SP. Supplementary eye field contrasted with the frontal eye field during acquisition of conditional oculomotor associations. J Neurophysiol 1995b;73:1122–1134. [PubMed: 7608759]

Chen LL, Wise SP. Evolution of directional preferences in the supplementary eye field during acquisition of conditional oculomotor associations. J Neurosci 1996;16:3067–3081. [PubMed: 8622136]

Corrado GS, Sugrue LP, Seung HS, Newsome WT. Linear Nonlinear Poisson models of primate choice dynamics. J Exp Anal Behav 2006;84:581–617. [PubMed: 16596981]

Curti, E.; Wang, XJ.; Fusi, S. Mechanisms for the formation of neural representations of abstract rules; CNS*2006 Proceedings; 2006;

Daw ND, Niv Y, Dayan P. Uncertainty based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 2005;8:1704–1711. [PubMed: 16286932]

Deco G, Rolls ET. Attention and working memory: a dynamical model of neuronal activity in the prefrontal cortex. Eur J Neurosci 2003;18:2374–2390. [PubMed: 14622200]

Deco G, Rolls ET. Synaptic and spiking dynamics underlying reward reversal in the orbitofrontal cortex. Cereb Cortex 2005;15:15–30. [PubMed: 15238449]

Fusi S. Hebbian spike driven synaptic plasticity for learning patterns of mean firing rates. Biol Cybern 2002;87:459–470. [PubMed: 12461635]

Fusi S, Drew PJ, Abbott LF. Cascade models of synaptically stored memories. Neuron 2005;45:599–611. [PubMed: 15721245]

Herrnstein, RJ.; Rachlin, H.; Laibson, DI. The Matching Law: Papers in Psychology and Economics. Harvard UP.: 1997.

Huang YY, Simpson E, Kellendonk C, Kandel ER. Genetic evidence for the bidirectional modulation of synaptic plasticity in the prefrontal cortex by D1 receptors. Proc Natl Acad Sci U S A 2004;101:3236–3241. [PubMed: 14981263]

Kennerley SW, Walton ME, Behrens TEJ, Buckley MJ, Rush-worth MFS. Optimal decision making and the anterior cingulate cortex. Nat Neurosci 2006;9:940–947. Comparative Study. [PubMed: 16783368]

Loewenstein Y, Seung HS. Operant matching is a generic outcome of synaptic plasticity based on the covariance between reward and neural activity. Proc Natl Acad Sci U S A 2006;103:15224–15229. [PubMed: 17008410]

Machon, A.; Pasupathy, A.; Histed, M.; Miller, E. Society for Neuroscience Abstracts; 2006. Learning-related changes in the caudate nucleus (cd) and frontal eye fields (fef) during a visuomotor task..

Mansouri FA, Tanaka K. Behavioral evidence for working memory of sensory dimension in macaque monkeys. Behav Brain Res 2002;136:415–426. [PubMed: 12429403]

Matsumoto K, Suzuki W, Tanaka K. Neuronal correlates of goal-based motor selection in the prefrontal cortex. Science 2003;301:229–232. [PubMed: 12855813]

Meyer, PL. Introductory probability and statistical applications. Addison-Wesley; Reading, MA: 1965.

Miyashita Y. Inferior temporal cortex: where visual perception meets memory. Ann. Rev. Neurosci 1993;16:245–263. [PubMed: 8460893]

Murray EA, Bussey TJ, Wise SP. Role of prefrontal cortex in a network for arbitrary visuomotor mapping. Exp Brain Res 2000;133:114–129. [PubMed: 10933216]

Origlia N, Kuczewski N, Aztiria E, Gautam D, Wess J, Domenici L. Muscarinic acetylcholine receptor knockout mice show distinct synaptic plasticity impairments in the visual cortex. J Physiol 2006;577:829–840. [PubMed: 17023506]

Otani S, Daniel H, Roisin M-P, Crepel F. Dopaminergic modulation of long-term synaptic plasticity in rat prefrontal neurons. Cereb Cortex 2003;13:1251–1256. [PubMed: 14576216]

Parisi G. A memory which forgets. J. Phys. A: Math. Gen 1986;19:L617.

Passingham, R. The Frontal Lobes and Voluntary Action. Oxford University Press; Oxford: 1993.

Pasupathy A, Miller EK. different time courses of learning-related activity in the prefrontal cortex and striatum. Nature 2005;433:873–876. [PubMed: 15729344]

Petrides M. Motor conditional associative-learning after selective prefrontal lesions in the monkey. Behav Brain Res 1982;5:407–413. [PubMed: 7126320]

Petrides M. Deficits on conditional associative learning tasks after frontal and temporal-lobe lesions in man. Neuropsychologia 1985;23:601–614. [PubMed: 4058706]

Renart A, Song P, Wang X-J. Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. Neuron 2003;38:473–485. [PubMed: 12741993]

Reynolds JN, Hyland BI, Wickens JR. A cellular mechanism of reward related-learning. Nature 2001;413:67–70. [PubMed: 11544526]

Reynolds JNJ, Wickens JR. Dopamine-dependent plasticity of corticostriatal synapses. Neural Netw 2002;15:507–521. [PubMed: 12371508]

Salinas E. Context dependent selection of visuomotor maps. BMC Neurosci 2004;5:47. [PubMed: 15563737]

Schultz W. Multiple reward signals in the brain. Nat Rev Neurosci 2000;1:199–207. [PubMed: 11257908]

Schweighofer N, Doya K. Meta-learning in reinforcement learning. Neural Networks 2003;16:5–9. [PubMed: 12576101]

Seung HS. Learning in spiking neural networks by reinforcement of stochastic synaptic transmission. Neuron 2003;40:1063–1073. [PubMed: 14687542]

Soltani A, Lee D, Wang X-J. Neural mechanism for stochastic behaviour during a competitive game. Neural Network 2006;19:1075–1090. Comparative Study.

Soltani A, Wang X-J. A biophysically-based neural model of matching law behavior: melioration by stochastic synapses. J. Neurosci 2006;26:3731–3744. [PubMed: 16597727]

Sugrue LP, Corrado GC, Newsome WT. Matching behavior and representation of value in parietal cortex. Science 2004;304:1782–1787. [PubMed: 15205529]

Suri RE, Schultz W. A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task. Neuroscience 1999;91:871–890. [PubMed: 10391468]

Sutton, R.; Barto, A. Reinforcement learning: an introduction. MIT Press; Cambridge, MA.: 1998.

Tanaka K. Inferotemporal cortex and object vision. Annu Rev Neurosci 1996;19:109–139. [PubMed: 8833438]

Tanaka SC, Doya K, Okada G, Ueda K, Okamoto Y, Yamawaki S. Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. Nat Neurosci 2004;7:887–893. [PubMed: 15235607]

Turrigiano GG. Homeostatic plasticity in neuronal networks: the more things change, the more they stay the same. Trends Neurosci 1999;22:221–227. [PubMed: 10322495]

Wallis JD, Anderson KC, Miller EK. Single neurons in prefrontal cortex encode abstract rules. Nature 2001;411:953–956. [PubMed: 11418860]

Wang X-J. Probabilistic decision making by slow reverberation in cortical circuits. Neuron 2002;36:955–968. [PubMed: 12467598]

White IM, Wise SP. Rule dependent neuronal activity in the prefrontal cortex. Exp Brain Res 1999;126:315–335. [PubMed: 10382618]

Williams R. Simple statistical gradient-following algorithms learning by stochastic hill climbing on discounted reward. Machine Learning 1992;8:229–256.

Wirth S, Yanike M, Frank LM, Smith AC, Brown EN, Suzuki WA. Single neurons in the monkey hippocampus and learning of new associations. Science 2003;300:1578–1581. [PubMed: 12791995]

Wixted JT, Ebbesen EB. Genuine power curves in forgetting: a quantitative analysis of individual subject forgetting functions. Mem Cognit 1997;25:731–739.

Yu AJ, Dayan P. Uncertainty, neuromodulation, and attention. Neuron 2005;46:681–692. [PubMed: 15944135]

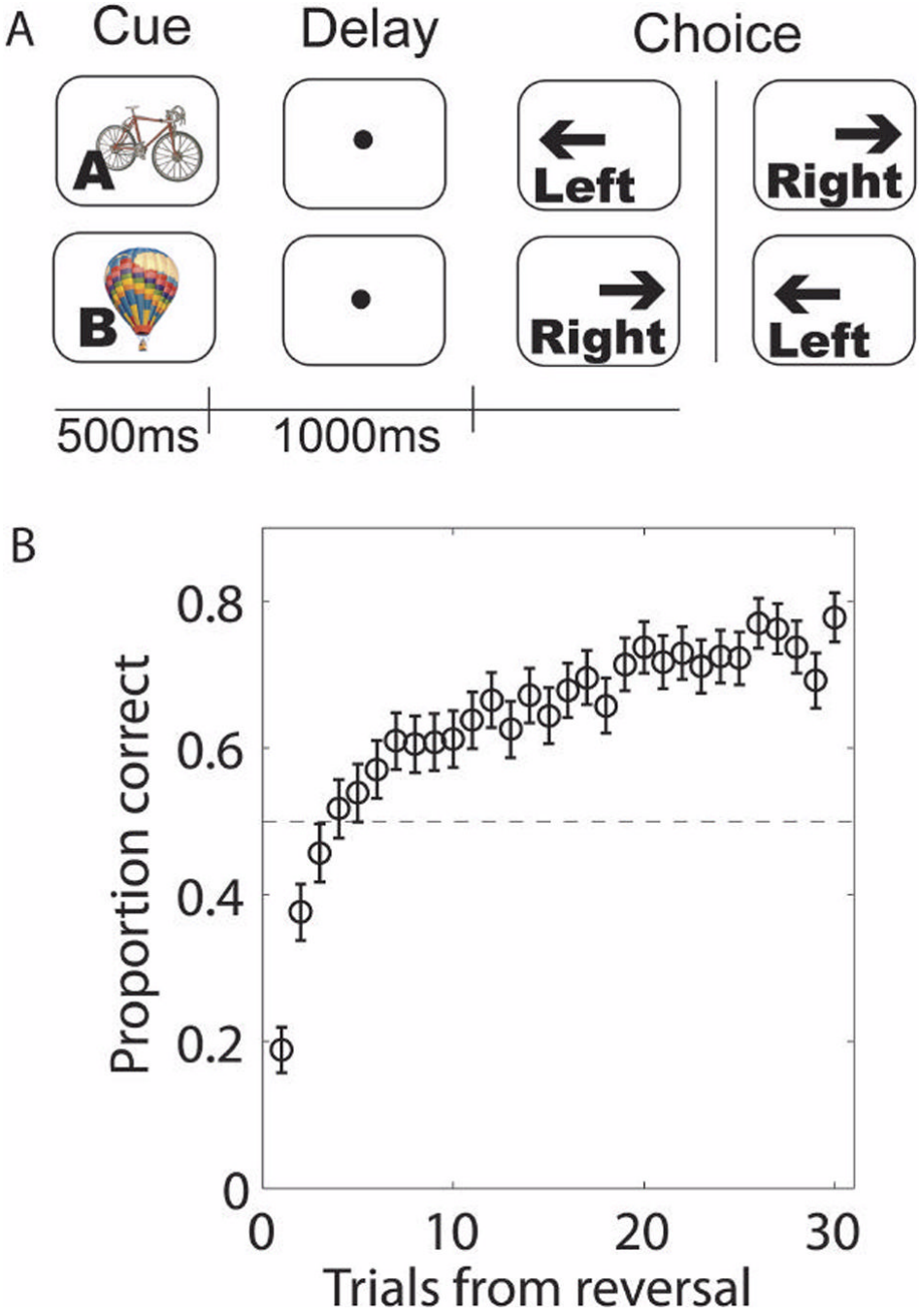

Figure 1.

Visuo motor association experiment. A) Task protocol: the monkey learns to associate four stimuli either with a left or right saccadic movement. The associations for two stimuli are reversed at unpredictable times and without explicit cues. For the other two stimuli (not shown) the associations are always the same (i.e. non reversing). B) The proportion of correct responses, averaged across all the blocks, is plotted against the number of trials from the time of reversal. Initially the monkey keeps responding according to the previously rewarded associations and makes the greatest number of mistakes. He forgets quickly (2-3 trials), whereupon performance rises to chance level (50%). The new associations are learned slowly (15-20 trials).

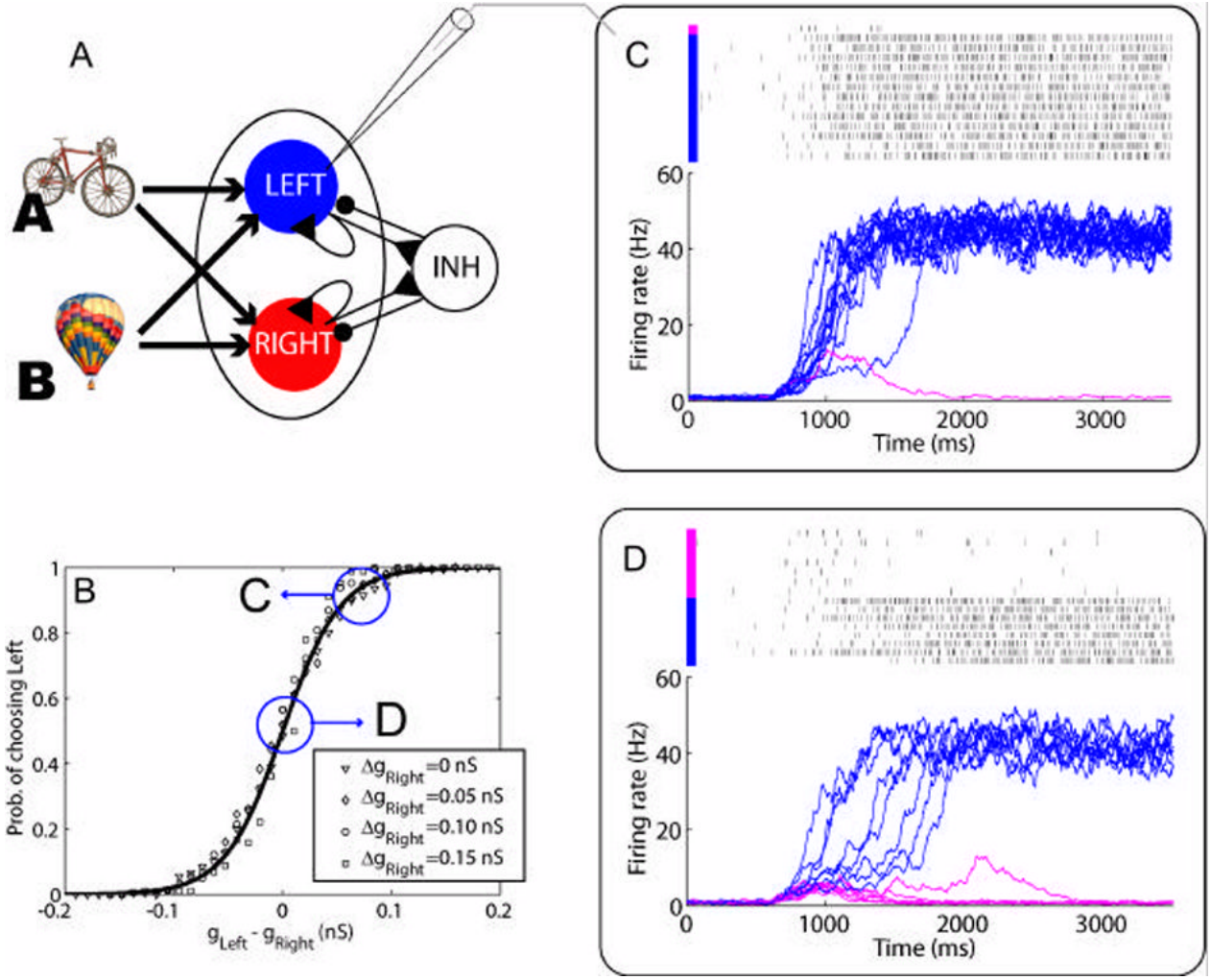

Figure 2.

The decision making network behavior. A) The architecture of the simulated decision making neural network: the network includes two excitatory subpopulations selective to the intended saccadic movements. The two populations compete through a group of inhibitory neurons. Visual stimuli activate the excitatory external inputs indicated by the black arrows. B) probability that Left wins over Right, as a function of the difference in the average external synaptic conductances. Different symbols correspond to different g_{Right} values (g_{Right} is 4.8nS plus the Δg_{Right} reported in the inset). The probability, computed by running full simulations of integrate and fire neurons (200 trials for each data point) is well described by a sigmoidal function (black line). The total mean external conductance $g_{Left} + g_{Right}$ for the points corresponding to $\Delta g_{Right} = 0$ (triangles) is 9.6nS. C,D) Raster plots for a single model neuron selective for Left, in 14 different trials. Blue: trials in which Left wins. Magenta: trials in which Left loses and Right is chosen. The parameters characterizing the statistics of the noisy input synaptic conductances are the same for all the trials shown in each of the two panels. different traces correspond to different realizations of the noisy inputs. C) a larger synaptic input g_{Left} than g_{Right} from the same stimulus (bicycle) makes Left choice more probable. D) with perfected balanced inputs Left is selected in half of the trials

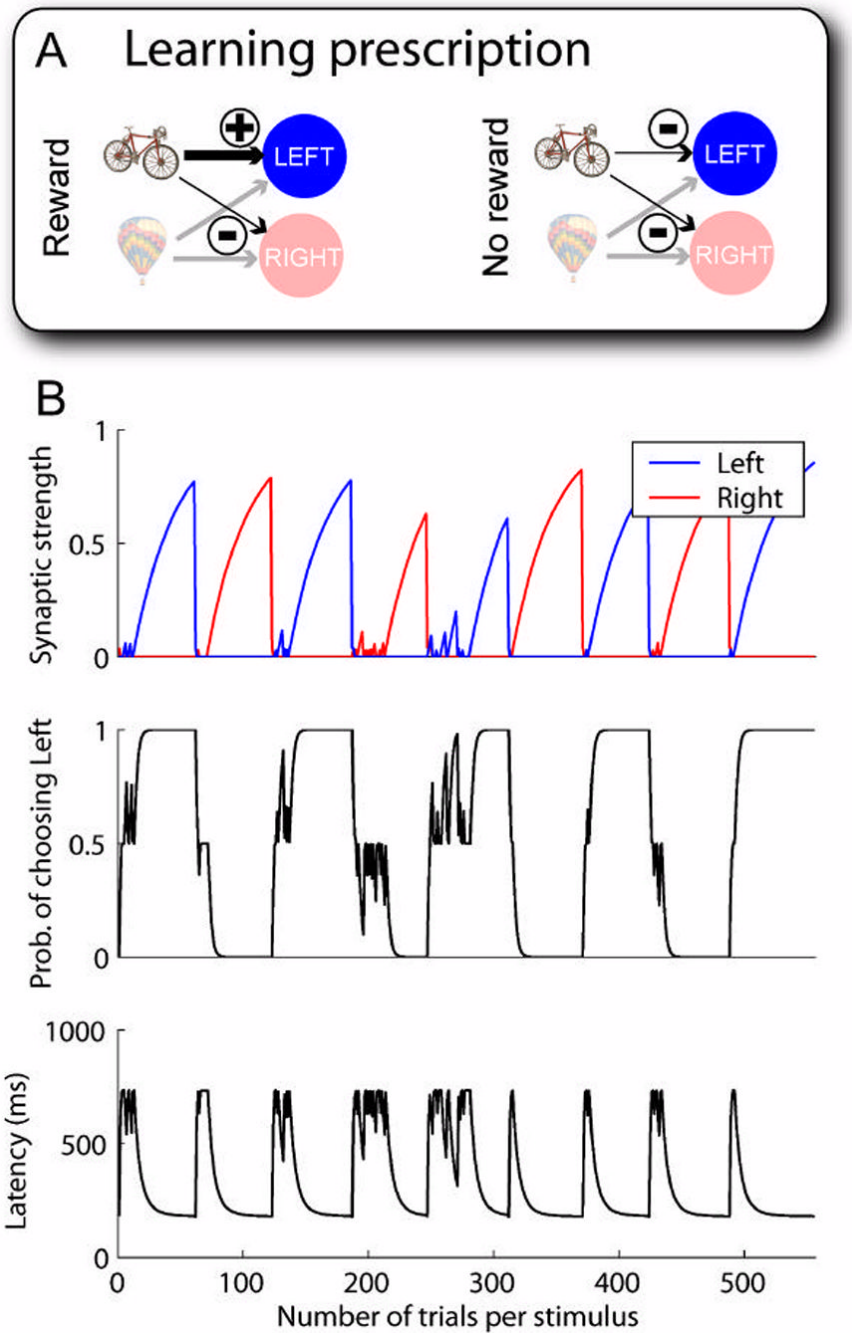

Figure 3.

Learning rewarding cue-response associations. A) learning scheme for the model when one stimulus is presented (say the bicycle) and Left neural population wins. If the association (bicycle-Left) is correct, the response leads to reward. In such a case the input to cells selective for Left is strengthened and the one to those selective for Right is weakened. If the association is incorrect (bicycle-right) and no reward is delivered, then both synaptic inputs are depressed and quickly brought to their minimal values which are assumed to be equal (symmetric configuration). B) simulation of the learning process for several blocks of trials in which the associations are reversed (each block has a random length between 60 and 70 trials). Top: synaptic strengths of input from a given stimulus to Left (blue) and to Right (red) neural pools

are plotted as a function of the trial number. Middle: the corresponding probability of choosing Left. Bottom: the latency to the half maximum selectivity which measures the speed of selecting a choice by decision neurons within a trial.

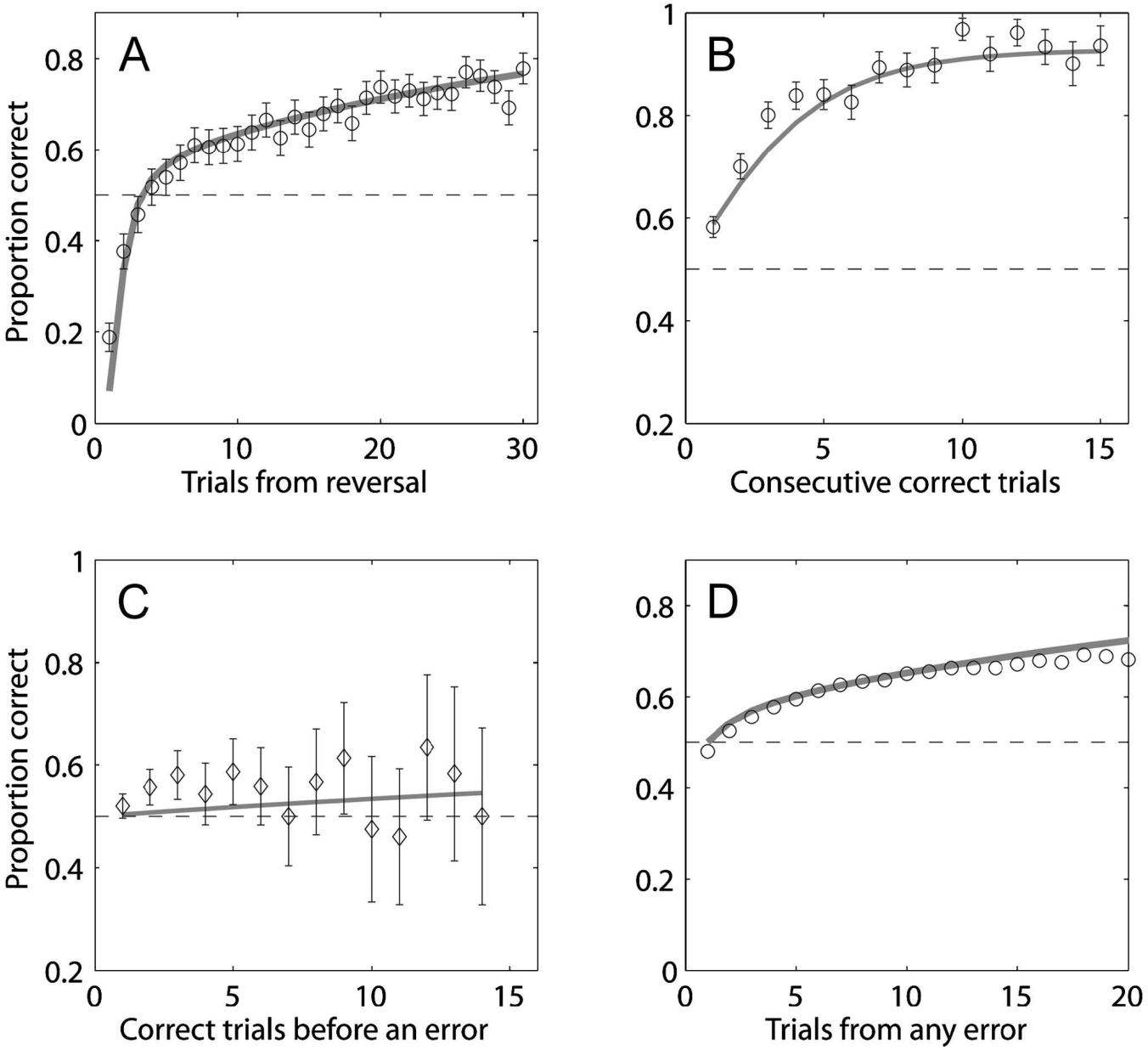

Figure 4.

The model reproduces quantitatively salient behavioral observations and makes predictions which are verified by analysis of monkey data. A) performance vs number of trials after the reversal of the associations for the model simulations (gray solid line) and for the experimental data (black points). B) the performance in a trial after n correct trials for the experimental data (with symbols) and simulation (gray line without symbols) C) Every mistake resets the monkey's performance to chance level: the probability of correct response in a trial following a single error is plotted as a function of the number of consecutive correct trials which precede the mistake. The performance is close to chance level, regardless of the length of the previous sequence of consecutively correct trials. The errors considered for the analysis can occur at any time within a block, and not necessarily immediately after reversal. D) performance vs number of trials after every error observed in the experiment (empty circles) compared to the same performance predicted by the model. The errorsbars for the datapoints are negligible.

The learning curve after reversal is very similar, indicating that the monkey relearns the associations in the same way, whether the error was caused by a reversal or by other reasons.

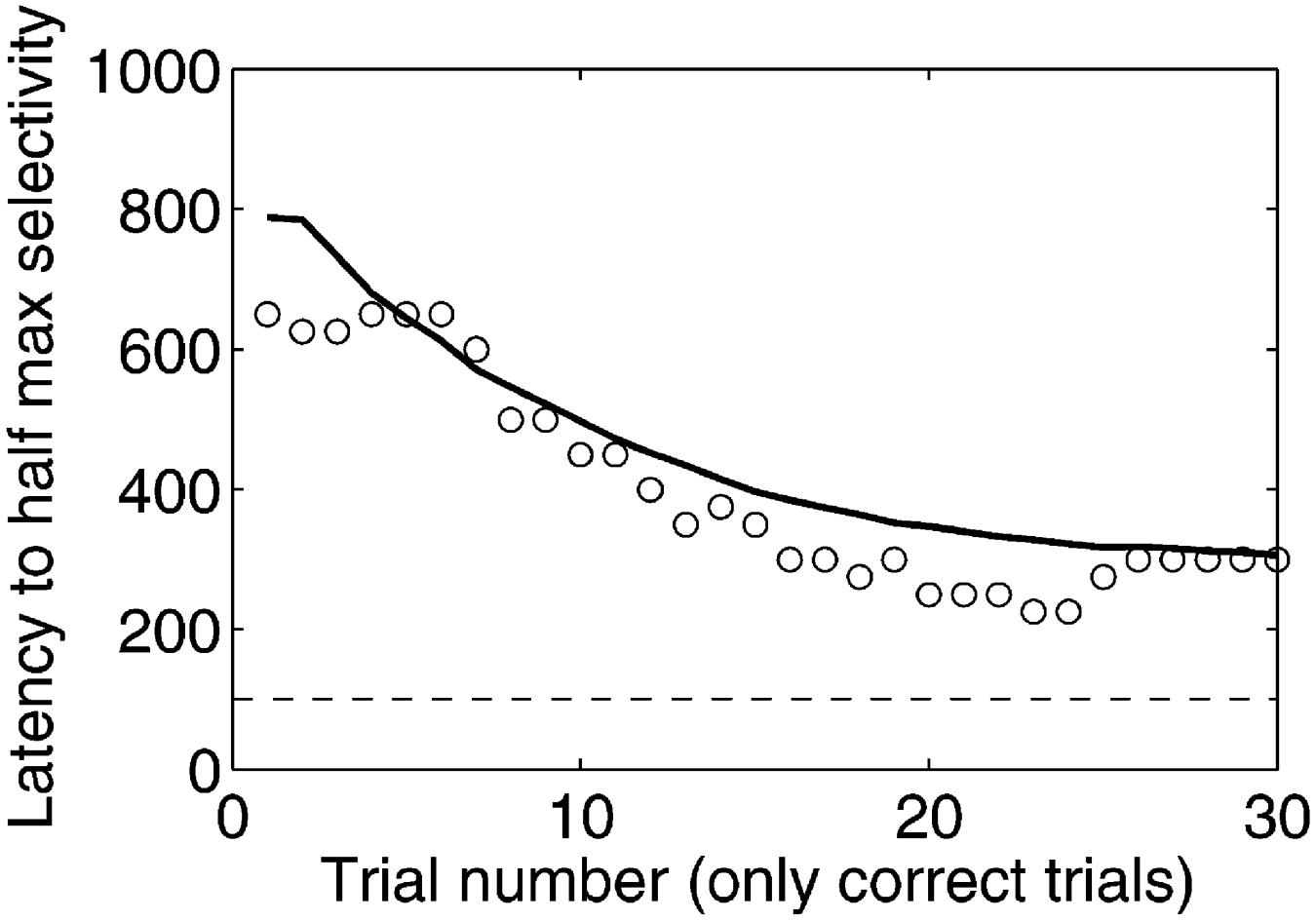

Figure 5.
Latency to the half maximum selectivity in simulations of the full network model of spiking neurons (solid line) and in the recorded prefrontal cells in the experiment (circles). Both are plotted as a function of the number of correct trials from the reversal of associations. The time is measured from the cue onset. The dashed line indicates the presumed latency (100ms) for a sensory signal to arrive to the recorded neuron.

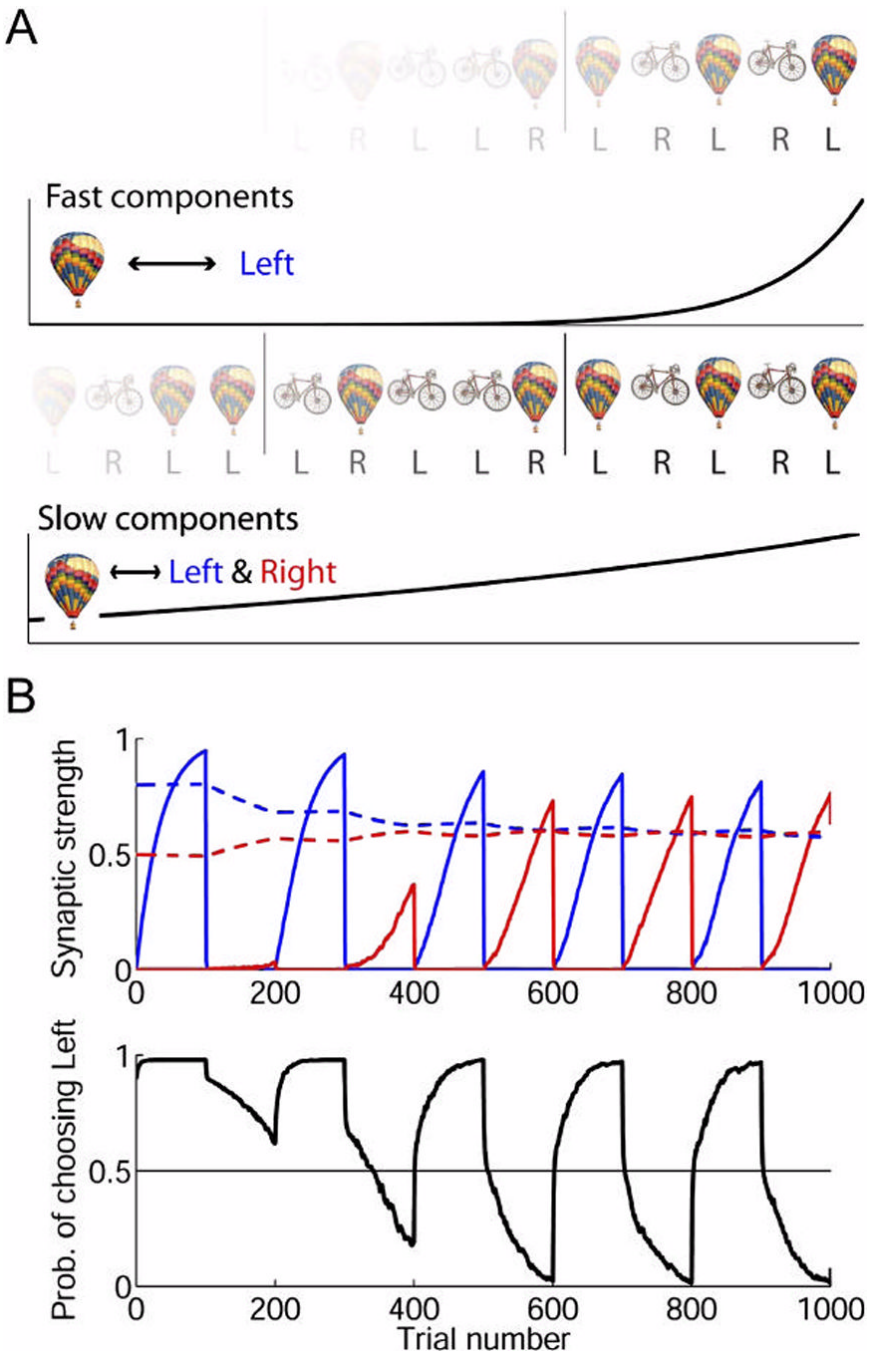

Figure 6.

Learning to decide probabilistically in the model. A) the role of fast and slow components of learning. The memory trace of past events (solid black line) decays exponentially with age (on the horizontal axis). The memorized events are schematically indicated by the visual stimulus and the corresponding rewarded response. Most recent memories are vivid, while the old ones fade away. The time constant of the exponential decay depends on the learning rate: components with high learning rates (fast components) forget quickly and typically remember only the associations within the last block of trials, whereas low learning rates (slow components) can produce memories which span more than one block (different blocks are separated by vertical lines). Hence the fast components of one stimulus (say the balloon) are

at any particular moment associated with only one saccade (Left in this example). For the slow components the same stimulus is associated with both saccadic movements. For example the balloon is remembered to be associated with both Left (in the present block), and Right (in the previous block). B) balanced network realized by a two component (fast and slow) learning process. Dashed lines: the slow components (blue and red: synaptic inputs to the Left and Right neural pools). Solid lines: the fast components. The simulation starts with a situation in which the difference in the slow components bias the response towards Left. After a few blocks in which the associations are reversed, the slow components balance each other. When the fast components are reset (as after a reversal), the model responds according to the configuration of the slow components which are now balanced, leading to a probabilistic choice behavior.

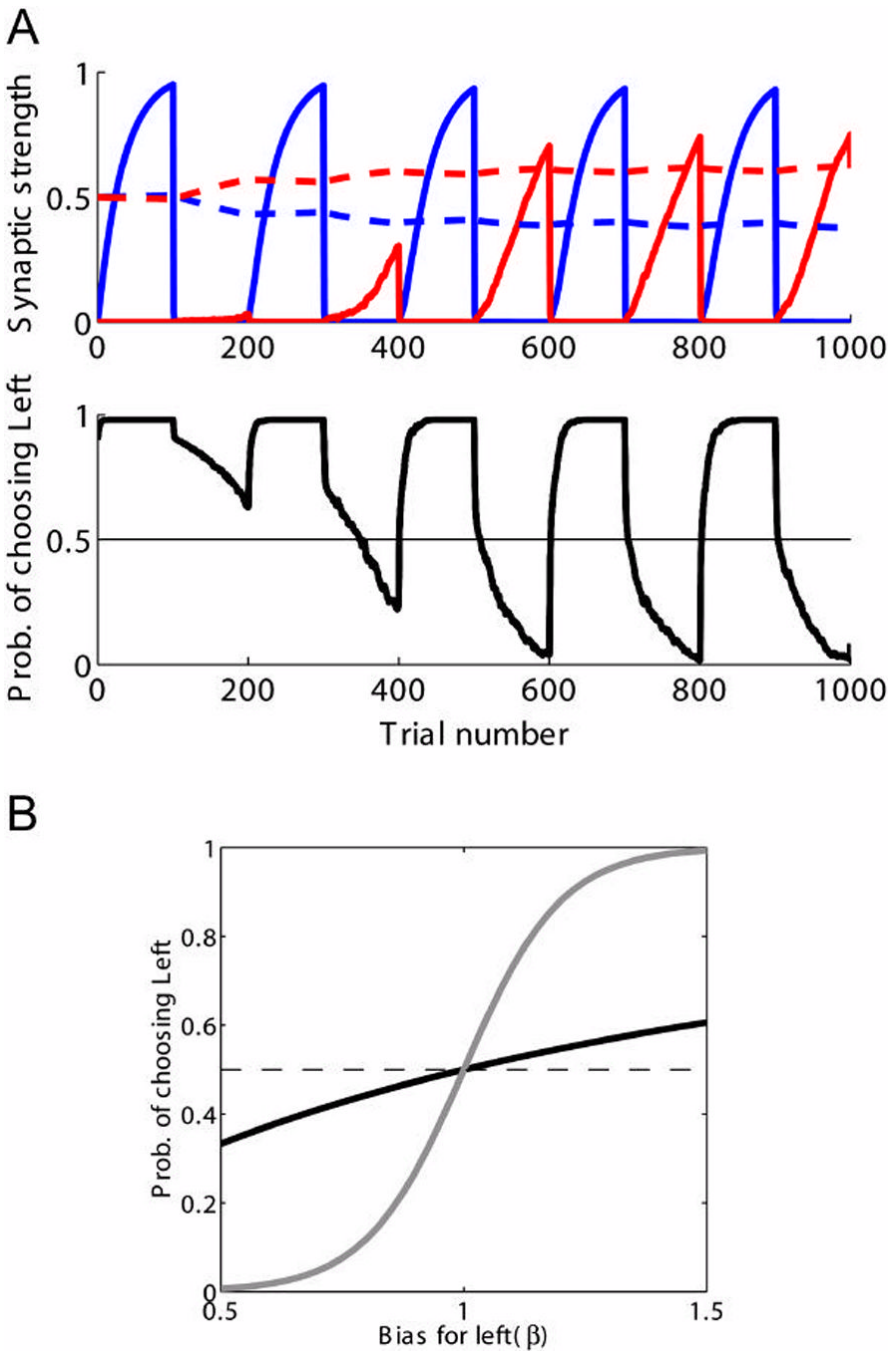

Figure 7.

Learning to behave randomly without fine-tuning. A) simulation of a network which has a preference for Left (e.g. because of a larger number of afferents to neurons selective for Left than those selective for Right). Top: fast and slow components of the synaptic inputs as a function of the number of trials, Bottom: the corresponding probability of choosing Left. The color code is the same as in Fig. 6. The strength of all the slow and fast components of input synaptic conductances are initially balanced, but the network chooses Left all the time because of the bias caused by heterogeneities. After a few blocks in which the two responses have exhibited an equal reward history, the slow component of the input to the Right neural pool (dashed red) becomes gradually larger than that to the Left neural pool (dashed blue).

Eventually the probability of choosing Left undergoes all association reversals, and any single mistake resetting the fast components would bring back to a balanced configuration (not shown). A small asymmetry still remains after learning (the speed of learning is slightly higher for the response Left) because the bias towards left is strong (the biasing parameter $\beta = 1.6$, see Experimental Procedure). B) The probability of choosing left vs the bias towards Left ($\beta = 1$ corresponds to no bias) is plotted before (gray) and after (black) learning. The black curve is much closer to the chance level (dashed line) (encoding the statistics of reward across different blocks) than the gray curve.

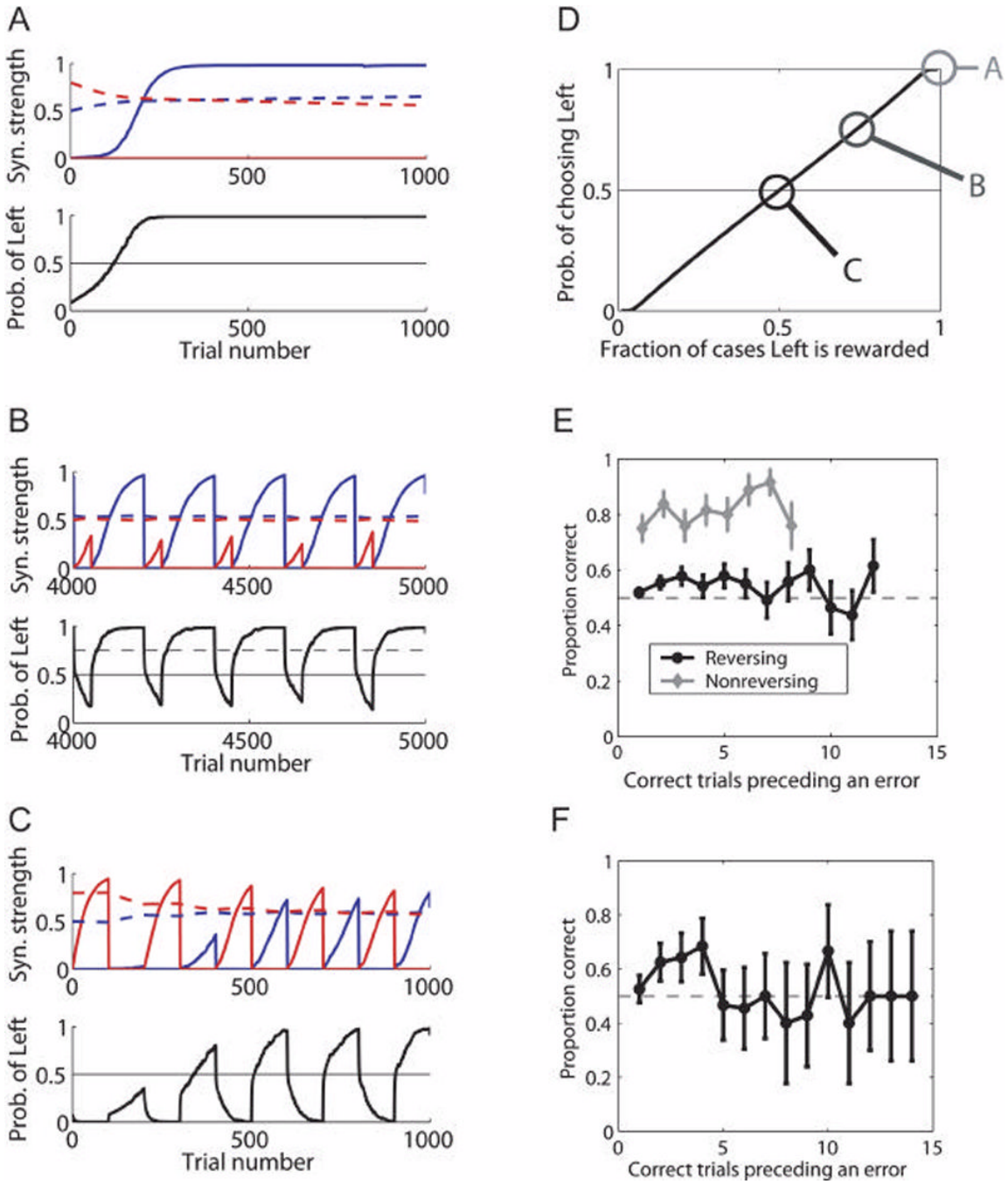

Figure 8.

The slow learning components encode reward history over long timescales. A,B,C) simulations of the learning dynamics of fast and slow components as described in Fig. 6 for three different reward statistics: A) the stimulus that we consider is always associated with Left, on long and short timescales (Left has probability 1 of being rewarded). This is the situation for the stimuli whose associations are never reversed. Both the fast and slow components strongly bias Left, no matter what is the initial condition. In this situation, single errors which reset the fast components, would not lead to chance level performance, because the slow components would still bias the choice for Left. B) the blocks in which the stimulus is associated with Left are longer than the blocks in which it is associated with Right (Left has probability 0.75 of being

rewarded when many blocks are considered). The probability of choosing Left (black solid line) reflects this statistics and it is shifted above the chance level (the simulation show the behavior after a large number of trials, when the slow components are at equilibrium). When the fast components reset immediately after a reversal, the decision network chooses Left with probability 0.75 (instead of 0.5). C) balanced statistics: Left and Right are equally probable correct associations across many blocks of trials. Any mistake should lead to random behavior and the two saccades would be equally likely. This is what happens in the experiment for the stimuli whose associations are reversed. D) summary of model behavior for different reward histories. The probability of choosing Left (averaged across many blocks o trials) matches the probability of Left being rewarded for a particular stimulus. E) behavioral data: the performance after a sequence of n correct trials followed by one error is plotted against n for the stimuli whose associations are reversed (black, same data as in Fig. 4C) and for the stimuli which are always consistently associated with one saccade (gray). Data points are shown only when the instances (number of trials) are more than 20. F) same as E, but restricted to the first blocks of each session (only stimuli whose associations are reversed are included). Although the statistics is poorer, the reset effect is evident also in this case when the stimuli are novel. See the Supplemental Material for more details.